UDC 551.577.2:551.577.36:551.501.45

A GENERALIZED PROBABILITY MODEL FOR SEQUENCES OF WET AND DRY DAYS

J. R. GREEN
University of Liverpool, England

ABSTRACT

A generalization of the two main probability models used to describe runs of wet days or of dry days is given. The new model depends on only two parameters and is shown to fit a large proportion of available sets of data.

1. INTRODUCTION

For some time it has been known that sequences of wet (or of dry) days exhibit persistence and do not conform to a constant probability (Bernoulli trials) model, whereby the probability of a wet day is independent of the number of immediately preceding wet or dry days. For example see Newnham (1916), the discussion by Brooks and Carruthers (1953) in chapter 16 of their book, and other references on this theme by Weiss (1964). Various alternative models have been suggested, the most popular of which has been a simple Markov model (see, for example, Chatfield 1966, Cooke 1953, Feyerherm and Bark 1965, Gabriel and Neumann 1962, Longley 1953, Weiss 1964). Certain models derived from particular continuous-time models were considered by Green (1964, 1965, 1967) and shown to be equivalent to the simple Markov model in most cases. Williams (1952) successfully fitted a logarithmic series to runs of wet days and runs of dry days at Harpenden, England. Cooke (1953) fitted the same type of series to rainfall records at Moncton, New Brunswick. Also Brooks and Carruthers (1953) put forward a modification of the Markov model whereby the probability of a day following a wet day is a constant from the third wet day onward, but this is different from the probability that the second day is wet. They successfully fitted this model to Newnham's data (1916) on wet days at Kew.

Wiser (1965) has also examined several modifications of the simple Markov model (using a three-urn illustration) and has shown how these fit more sets of data than the unmodified form. We feel that the new model proposed in this paper is simpler than Wiser's modifications, and it certainly has a high success rate in application to different sets of data.

2. RELATION OF THE NEW MODEL TO SOME EARLIER MODELS

The probabilities of runs of 1, 2, . . . , r, . . . days of wet (or dry) days according to certain models are consecutively proportional to the quantities shown (where q and a are parameters):

Model 1—Williams' log series, $q, q^2/2, \ldots, q^r/r, \ldots$, with normalizing constant $-1/\log(1-q)$.

Model 2—Markov chain, $q, q^2, \ldots, q^r, \ldots$, with normalizing constant $(1-q)/q$.

Model 3—new modified log model, $q/(1+a), q^2/(2+a), \ldots, q^r/(r+a), \ldots$, where a is a constant between 0 and infinity and the normalizing constant, c, such that $c\Sigma q^r/(r+a)=1$, is easily computable.

Incidentally, the Bernoulli model, like the Markov model, also has run probabilities which form a geometric series, but in this case the probability of a day being wet is independent of the states of the previous days, whereas in the Markov model the probability of a wet day *does* depend on the state of the preceding day.

Letting W represent a wet day, D a dry day, and $P(A|B)$ the probability of A given B, we have the probability of a run of r wet days (any r, the first wet day being given) is $P(W^{r-1}D|W)=P(W|W)\cdot P(W|W^2) \ldots P(W|W^{r-1})\times P(D|W^r)$. For the Bernoulli model,

$$P(W^{r-1}D|W)=P(W)^{r-1}P(D).$$

For model 2,

$$P(W^{r-1}D|W)=P(W|W)^{r-1}P(D|W).$$

In models 1 and 3, these probabilities are more simply and directly defined as functions of r such that

$$P(W^{r-1}D|W)=\begin{cases} q^r/r, & \text{for some } q \text{ (model 1)} \\ q^r/(r+a), & \text{for some } q, r \text{ (model 3).}\end{cases}$$

The new model 3 includes models 1 and 2, by a taking the values 0 and ∞, respectively. For a greater than 1, it is more convenient, for computational purposes, to write the model 3 probabilities as being proportional to $q/(1+a_1), q^2/(1+2a_1), \ldots, q^r/(1+ra_1), \ldots$. Here, a_1 is the reciprocal of a as used above, and so, for a greater than 1, a_1 lies between 0 and 1.

3. THE GOODNESS OF FIT OF DATA TO MODELS 1, 2, AND 3

Models 1 and 2 have been shown to fit a large proportion of the appropriate sets of data available, as demonstrated by table 1. The first 11 places in this table, down to Fort Worth, were discussed by Weiss (1964) and also partly by Green (1965) and Wiser (1965).

TABLE 1.—*Sets of data fitting models 1, 2, and 3 (according to the χ^2 test at 5 percent level)*

	Set of data		Model 1	Model 2	Neither 1 nor 2	Model 3	a^*	q^*
1	Montsouris (Besson 1924)	Wet			√	√	0.3	0.79
2	San Francisco	Dry			√	√	1.25	0.91
3	(Jorgensen 1949)	Wet		√		√	2.5	0.72
4	Harpenden	Dry	√					
5	(Williams 1952)	Wet			√	√	0.9	0.795
6	Montreal	Dry		√				
7	(Longley 1953)	Wet		√				
8	Moncton	Dry			√	X	∞	0.725
9	(Cooke 1953)	Wet			√	X	0.24	0.518
10	Storm area I	Dura.		√				
11	(Weiss 1964)	Int.		√				
12	Storm area II	Dura.		√		√	2.8	0.74
13	(Weiss 1964)	Int.		√		√	∞	0.42
14	Storm area III	Dura.		√		√	10	0.72
15	(Weiss 1964)	Int.		√		(√?)	∞	0.438
16	Storm area IV	Dura.		√				
17	(Weis 1964)	Int.		√				
18	Kansas City	Dry		√				
19	(Weiss 1964)	Wet		√				
20	Fort Worth	Dry		√				
21	(Weiss 1964)	Wet		√				
22	March	Dry	√			√	0	0.868
23	(Green, new data)	Wet			√	√	0.9	0.689
24	Aberdeen	Dry	√			√	0.28	0.747
25	(Newnham 1916)	Wet			√	X	5.6	0.770
26	Kew	Dry	√			√	0.09	0.83
27	(Newnham 1916)	Wet			√	√	0.93	0.75
28	Valencia	Dry	√			√	0	0.784
29	(Newnham 1916)	Wet			√	√	0.9	0.93
30	Greenwich	Dry	√			√	0.20	0.845
31	(Newnham 1916)	Wet			√	√	2.6	0.705
32	Kew	Dry	√					
33	(Chatfield 1966)	Wet		√				

√ = χ^2 test performed and nonsignificant result obtained.
X = χ^2 test performed and significant result obtained.

Although not the same kind of data as the rest which concern us here, Weiss' data concerning the duration of and intervals between storms in certain areas exhibit a similar probalistic behavior and were considered here also, and the relevant test results are included in table 1. It had previously been reported by the present author that the Markov model did not fit the data of areas II and III, whereas table 1 now indicates otherwise. Indeed there *is* a significant difference between the observed and computed values for each of these sets of data as shown in Weiss' table 2. However, the computed values for these sets, and the corresponding conditional probabilities, as shown in Weiss' table 2, were in error, and when the correct values are used the test results shown here in table 1 are obtained. However, we should mention that the fit of the model is nearly significantly bad (at the 5 percent level) for the storm durations of area II and intervals of area III; also the fit of model 3 is actually significantly bad in the latter case (as there is one degree of freedom less for chi-squared), although the fit of model 2 (a special case of model 3) is *just* not significant.

Williams' investigation of dry spells at Harpenden and Chatfield's investigation of wet and dry spells at Kew (which incidentally used later data than did the investigations of Newnham relating to Kew, 1958–65 as compared with 1901–10) demonstrated the successful fits of the models shown in table 1, but without performing goodness-of-fit significance tests. However, the fits were successfully tested, using chi-squared with 5 percent significance levels, by the present author. For the rest of the rows of table 1, wherever neither model 1 nor model 2 had been shown statistically, in previous published work, to fit the data, model 3 was fitted and the fit tested by a chi-squared goodness-of-fit test with 5 percent significance level. In some of these cases (as table 1 shows) it happens that model 1 or model 2 does fit the data, but this had not been shown in previous publications.

Actually Cooke (1953) had previously reported the Moncton data as fitting the log series for the dry runs, and the Markov model for the wet runs. These models may indeed be useful approximations to the rainfall behavior there, but the two fits were significantly bad by the chi-squared test at the 5 percent significance level, as indicated in table 1.

We see that, of the 33 sets of data here considered, seven fitted model 1, 16 fitted model 2, and seven of the remaining 10 sets fitted model 3. In all, 30 of the 33 sets fitted model 3, or 29, if one regards the data of the area III intervals as not fitting model 3. If in fact model 3 does apply in *all* these cases, then the probability of getting at least three results significant at the 5 percent level is 23.0 percent (the probability of at least four being significant is 8.6 percent). Thus our obtaining only three or four significant results out of 33 is quite consistent with the hypothesis that model 3 applies for all cases.

Since the data for March, England, have not been previously published, they are given here in table 2, together with the computed numbers of wet runs and numbers of dry runs of different lengths, according to model 3, for the period 1887–1918 (all seasons). It happens that in the case of the dry runs, the best model 3 fit is obtained by fitting the particular case, model 1.

Figures 1 and 2 also illustrate the measure of agreement between the data for March and the computed run distributions according to model 3.

4. FITTING MODEL 3

To estimate the parameters for model 3, whereby the probability of a run of r days is proportional to $q^r/(r+a)$,

TABLE 2.—*Observed and computed data for March, England*

Run length (days)	Observed no. wet runs	Computed no. wet runs	Observed no. dry runs	Computed no. dry runs	Tail of dry run distribution: Run length	Tail of dry run distribution: Observed no. runs
1	1102	1094.9	902	934.9		
2	480	494.3	417	405.7	19	2
3	277	253.2	246	234.8	20	2
4	129	138.9	157	152.8	21	1
5	76	79.5	111	106.1	22	2
6	52	46.8	97	76.8	23	2
7	20	28.2	54	57.1	24	1
8	12	17.2	41	43.4	25	1
9	15	10.7	24	33.5	26	1
10	6	6.7	32	26.1	27	0
11	7	4.2	25	20.6	28	2
12	2		13	16.4	29	1
13	2		12	13.2	30	2
14	-		12	10.6	31 or more	0
15	-	6 7.4	8	8.6		
16	1		0	7.0	Total	17
17	-		8	5.7		
18	1		5	4.7		
19 or more	-		17	23.0		
Total	2182	2182.0	2181	2181.0		

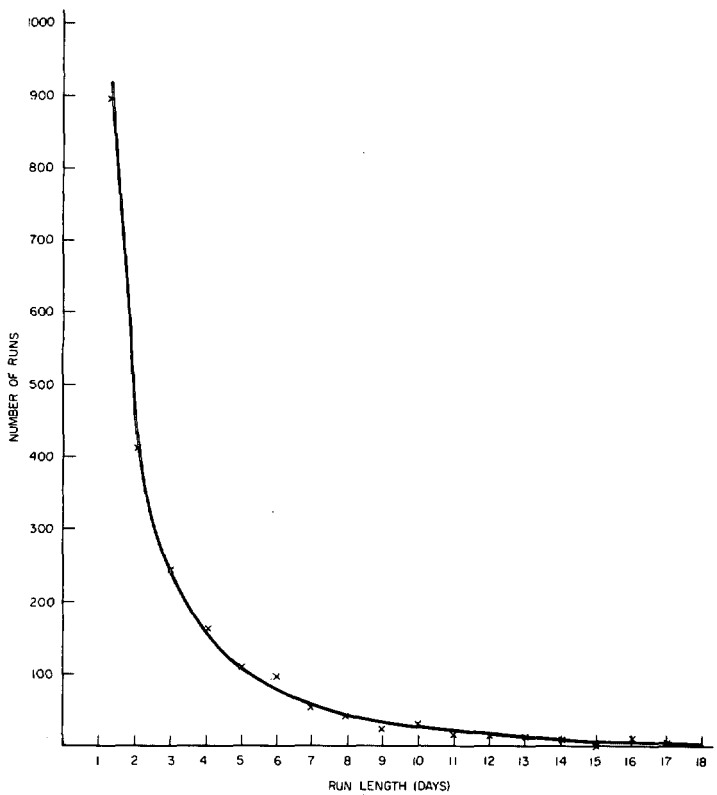

FIGURE 1.—Distribution of lengths of runs of dry days at March (England) for 1887–1918. Observed data, crosses; model 3 fit, full line.

it is appropriate to use either 1) the method of minimum chi-squared, or 2) the method of maximum likelihood. The two methods are asymptotically equivalent.

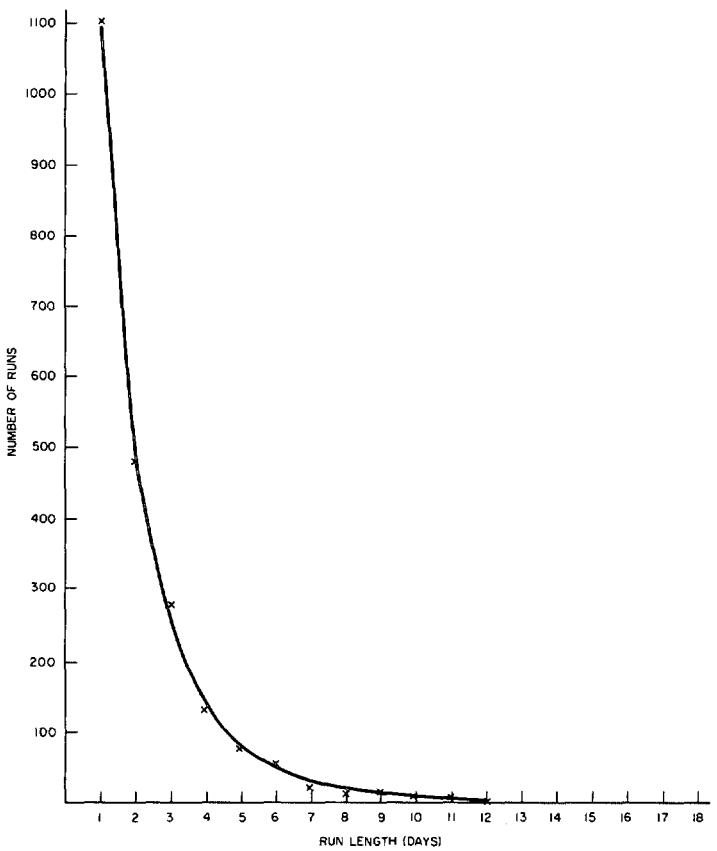

FIGURE 2.—Distribution of lengths of runs of wet days at March, England, for 1887–1918. Observed data, crosses; model 3 fit, full line.

1) The method of minimum chi-squared was used by the present author, by means of repeated application of a simple ALGOL program on an ICL KDF9 computer, using an Egdon operating system. This program calculates chi-squared at each point of a mesh of values of a and q for each set of data. The procedure can be repeated using a finer mesh in the area of the minimizing a and q values, for each set of data, and again if required. If approximate a and q values give a chi-squared value sufficiently small to be nonsignificant, then clearly the exact minimizing values will give a nonsignificant result. Each usage of the program required computer time of about 45–50 sec plus about 0.17 sec per a–q combination.

2) The method of maximum likelihood here produces two equations, which are satisfied by the estimates of a and q, namely (with r as the length of a run)

sample average, $\overline{(r+a)^{-1}}$=expected value of $(r+a)^{-1}$

and

sample average, $\bar{r}$=expected value of r,

that is, respectively,

$$N^{-1}\Sigma O_r/(r+a)=c\Sigma q^r/(r+a)^2 \qquad (1)$$

and

$$N^{-1}\Sigma rO_r = c\Sigma rq^r/(r+a) \qquad (2)$$

where O_r=observed number of runs of length r, $N=\Sigma O_r$, and c is the normalizing constant, such that $c\Sigma q^r/(r+a)=1$.

Using $\mathcal{E}$ for "expected value of," the values of $\mathcal{E}(r+a)^{-1}$ and $\mathcal{E}r$ are not readily expressible by convenient formulas, though they may be calculated easily and tabulated for different values of a and q. However, equations (1) and (2) are not easy to solve, since a appears awkwardly on both sides of equation (1), and is needed to apply equation (2).

An iterative graphical method would appear to be the best way to obtain an approximate solution, but attempts by the author to do this have not been very successful.

REFERENCES

Besson, Louis, "Sur le probabilite de la Pluie," (On the Probability of Rain), *Comptes Rendus*, Tome 152, A 181, May 19, **1924**, pp. 1743–1745.

Brooks, C. E. P., and Carruthers, N., *Handbook of Statistical Methods in Meteorology*, Her Majesty's Stationery Office, London, **1953**, 413 pp.

Chatfield, C., "Wet and Dry Spells," *Weather*, Vol. 21, No. 9, Sept. **1966**, pp. 308–310.

Cooke, D. S., "The Duration of Wet and Dry Spells at Moncton, New Brunswick," *Quarterly Journal of the Royal Meteorological Society*, Vol. 79, No. 342, Oct. **1953**, pp. 536–538.

Feyerherm, A. M., and Bark, L. Dean, "Statistical Models for Persistent Precipitation Patterns," *Journal of Applied Meteorology*, Vol. 4, No. 3, June **1965**, pp. 320–328.

Gabriel, K. R., and Neumann, J., "A Markov Chain Model for Daily Rainfall Occurrence at Tel Aviv," *Quarterly Journal of the Royal Meteorological Society*, Vol. 88, No. 375, Jan. **1962**, pp. 90–95.

Green, J. R., "A Model for Rainfall Occurrence," *The Journal of the Royal Statistical Society*, Vol. 26, No. 2, **1964**, pp. 345–353.

Green, J. R., "Two Probability Models for Sequences of Wet or Dry Days," *Monthly Weather Review*, Vol. 93, No. 3, Mar. **1965**, pp. 155–156.

Green, J. R., "A Modified Model for Rainfall Occurrence," *The Journal of the Royal Statistical Society*, Vol. 29, No. 1, **1967**, pp. 151–153.

Jorgensen, Donald L., "Persistency of Rain and No-Rain Periods During the Winter at San Francisco," *Monthly Weather Review*, Vol. 77, No. 11, Nov. **1949**, pp. 303–307.

Longley, Richmond W., "The Length of Dry and Wet Periods," *Quarterly Journal of the Royal Meterological Society*, Vol. 79, No. 342, Oct. **1953**, pp. 520–527.

Newnham, E. V., "The Persistence of Wet and Dry Weather," *Quarterly Journal of the Royal Meteorological Society*, Vol. 42, No. 179, July **1916**, pp. 153–162.

Weiss, Leonard L., "Sequences of Wet or Dry Days Described by a Markov Chain Probability Model," *Monthly Weather Review*, Vol. 92, No. 4, Apr. **1964**, pp. 169–176.

Williams, C. B., "Sequences of Wet and Dry Days Considered in Relation to the Logarithmic Series," *Quarterly Journal of the Royal Meteorological Society*, Vol. 78, No. 335, Jan, **1952**. pp. 91–96

Wiser, E. M., "Modified Markov Probability Models of Sequences of Precipitation Events," *Monthly Weather Review*, Vol. 93, No. 8, Aug. **1965**, pp. 511–516.

[*Received August 5, 1969; revised October 2, 1969*]